

## **Project 2 Report**

**TOPIC: CUSTOMER SEGMENTATION USING RFM ANALYSIS**

**IE6400 19843 Foundations Data Analytics Engineering  
SEC 01 Fall 2023**

### **Group Number 7**

Aishwariya Alagesan (002818973)  
Arya Lokesh Gowda (002249418)  
Keerthana Balswamy (002839068)  
Deepana Dhakshinamurthy (002834708)  
Suja Ganesh Murugan (002839290)

## **Introduction:**

This retail sales analysis report presents a comprehensive overview of a dataset containing retail sales data. The report covers various aspects of the dataset, including data preprocessing, customer analysis, product analysis, time analysis, geographical analysis, customer behavior, returns and refunds, revenue analysis, RFM (Recency, Frequency, Monetary) analysis, segmentation, and marketing recommendations.

## **Data Preprocessing**

- **Initial Data Inspection:**

Upon loading the dataset, we conducted an initial inspection to understand its structure and characteristics. The dataset consists of information related to retail transactions, including columns for Invoice Number, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country.

- **Data Cleaning:**

Data cleaning involved handling missing values, ensuring data consistency, and addressing duplicates. Missing values were identified and removed from the dataset. This step ensures data quality and prevents inconsistencies.

- **Data Transformation:**

The 'InvoiceDate' column was converted to a datetime format, and the 'CustomerID' column was converted to 64-bit integers. Additionally, duplicates were removed from the dataset.

## **Data Overview**

### **Dataset Size:**

The cleaned dataset contains a total of 401604 rows and 10 columns.

### **Column Descriptions:**

The columns in the dataset include:

- 'InvoiceNo': Unique invoice number for each transaction.
- 'StockCode': Code identifying the product.
- 'Description': Description of the product.
- 'Quantity': The number of items purchased.
- 'InvoiceDate': Date and time of the transaction.
- 'UnitPrice': Price per unit of the product.
- 'CustomerID': Unique customer identifier.
- 'Country': The country in which the transaction took place.

## Time Period Covered:

The dataset covers a time from 12/01/2010 to 12/08/2011.

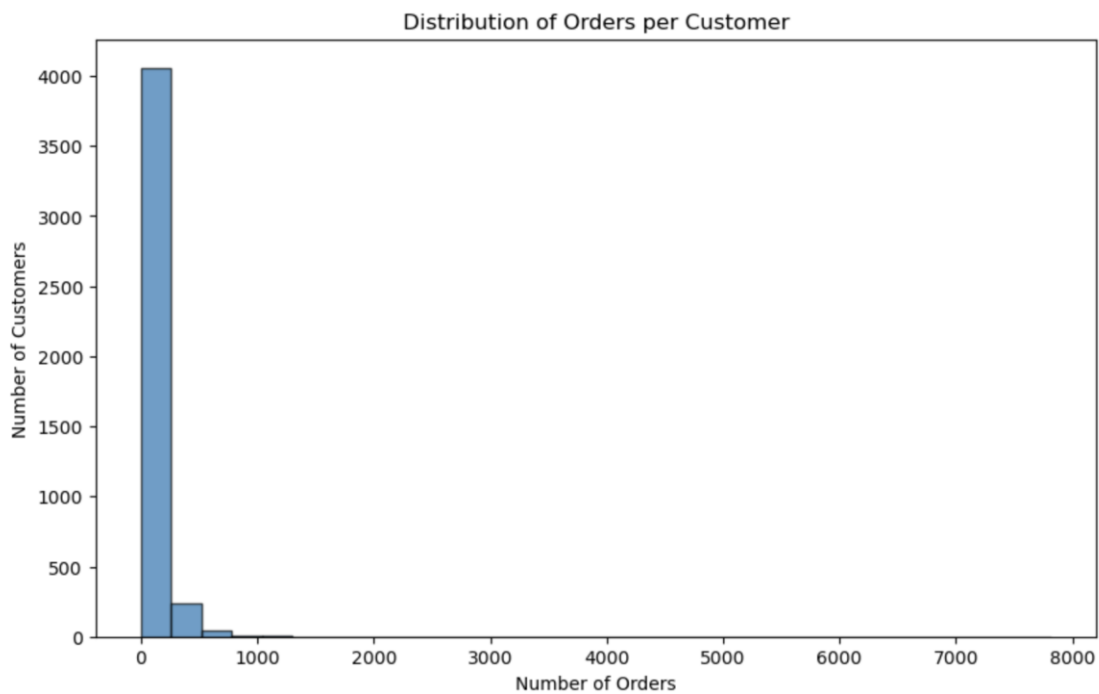
## Customer Analysis

- **Unique Customers:**

The dataset contains data for 4372 unique customers.

- **Distribution of Orders:**

The distribution of the number of orders per customer reveals:

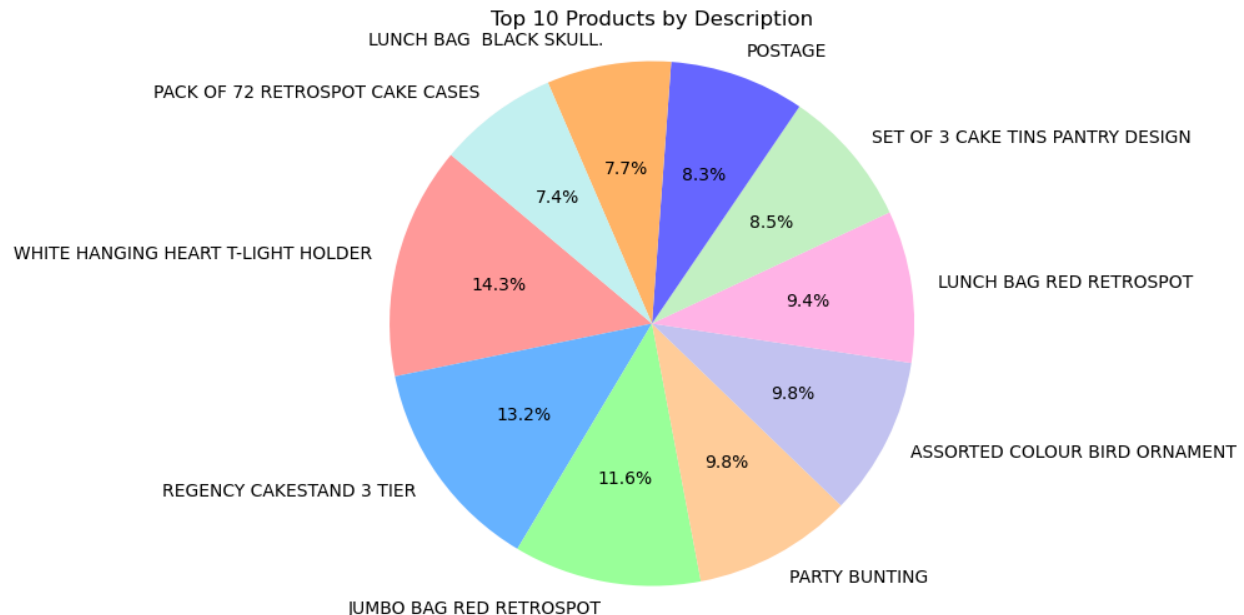


## Product Analysis

- **Top 10 Products:**

The top 10 most frequently purchased products are:

WHITE HANGING HEART T-LIGHT HOLDER	2058
REGENCY CAKESTAND 3 TIER	1894
JUMBO BAG RED RETROSPOT	1659
PARTY BUNTING	1409
ASSORTED COLOUR BIRD ORNAMENT	1405
LUNCH BAG RED RETROSPOT	1345
SET OF 3 CAKE TINS PANTRY DESIGN	1224
POSTAGE	1196
LUNCH BAG BLACK SKULL.	1099
PACK OF 72 RETROSPOT CAKE CASES	1062



- **Average Price of Products:**

The average price of products in the dataset is \$3.47406. Understanding the average price is essential for evaluating customer behavior and overall business profitability.

### Product Category with Highest Revenue:

Among the product categories, the category with the highest revenue is REGENCY CAKESTAND 3 TIER

A total revenue of \$132567.69 was made in this period.

## Time Analysis

- **Peak Order Times:**

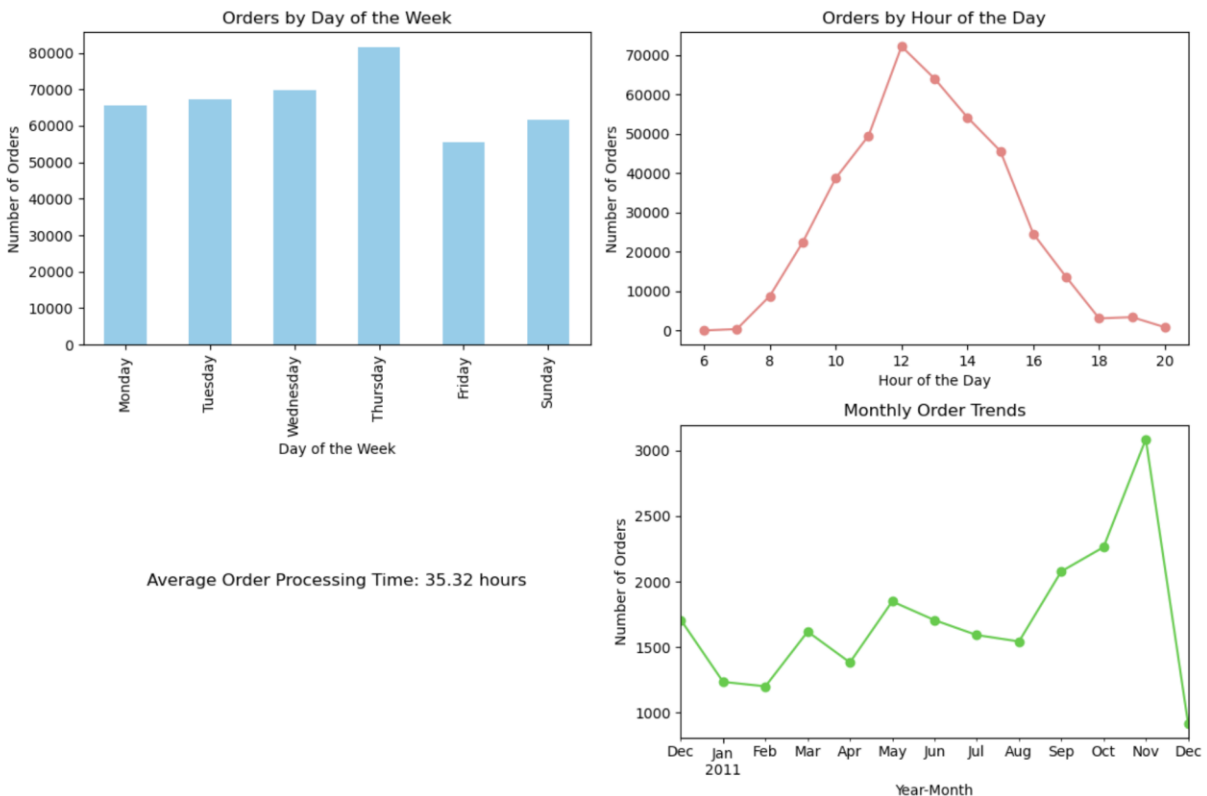
The analysis reveals that Thursday was when most orders are placed. This insight can inform inventory management and staffing decisions.

- **Average Order Processing Time:**

The average order processing time is 35.32 hrs. Understanding the order processing time is crucial for improving operational efficiency.

## Seasonal Trends:

Seasonal trends in the dataset are observed, with holiday season approaching a lot of purchases were recorded during the months of October-December. This understanding can guide marketing and inventory strategies.



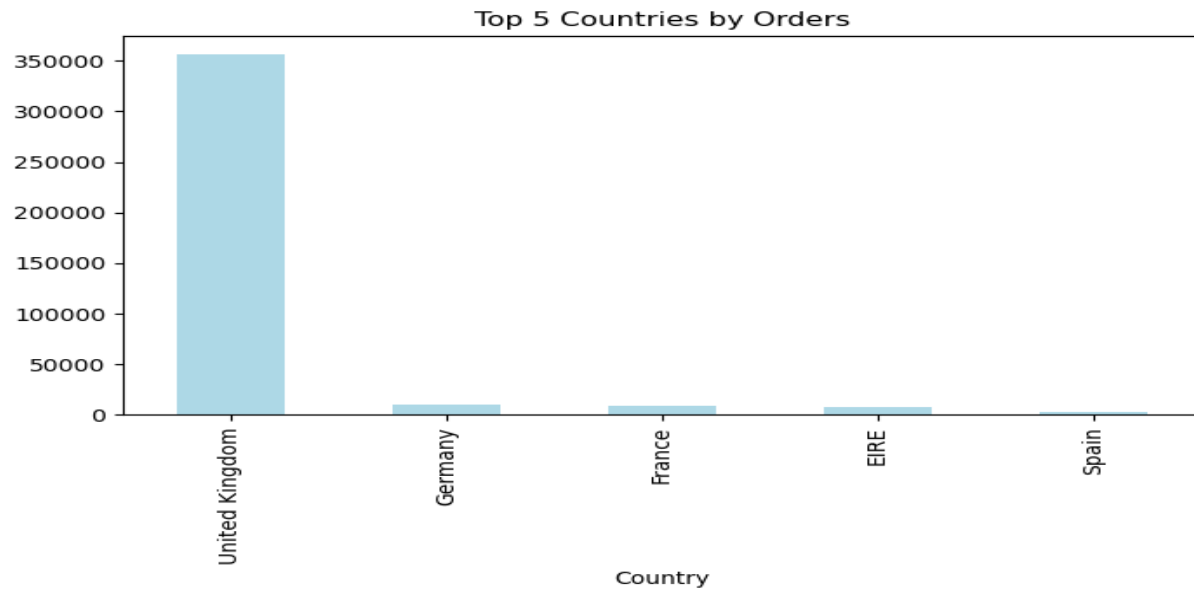
## Geographical Analysis

- **Top 5 Countries:**

The top 5 countries with the highest number of orders are:

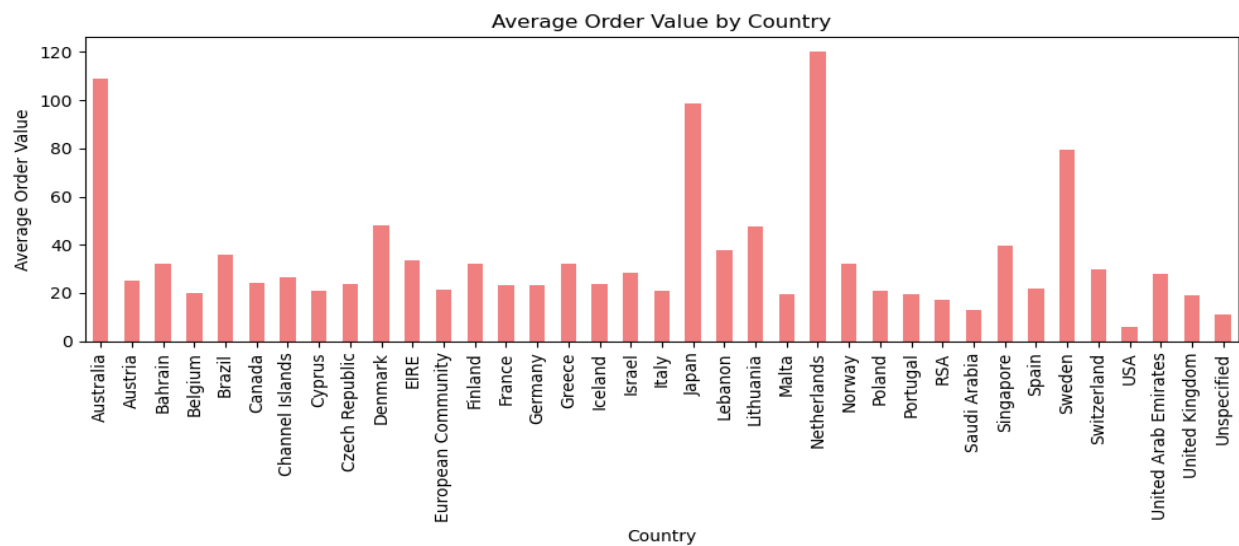
1. United Kingdom
2. Germany
3. France
4. EIRE
5. Spain

These countries contribute significantly to the business's international presence.



### Country-Customer Correlation:

There is a correlation between the customer's country and the average order value.



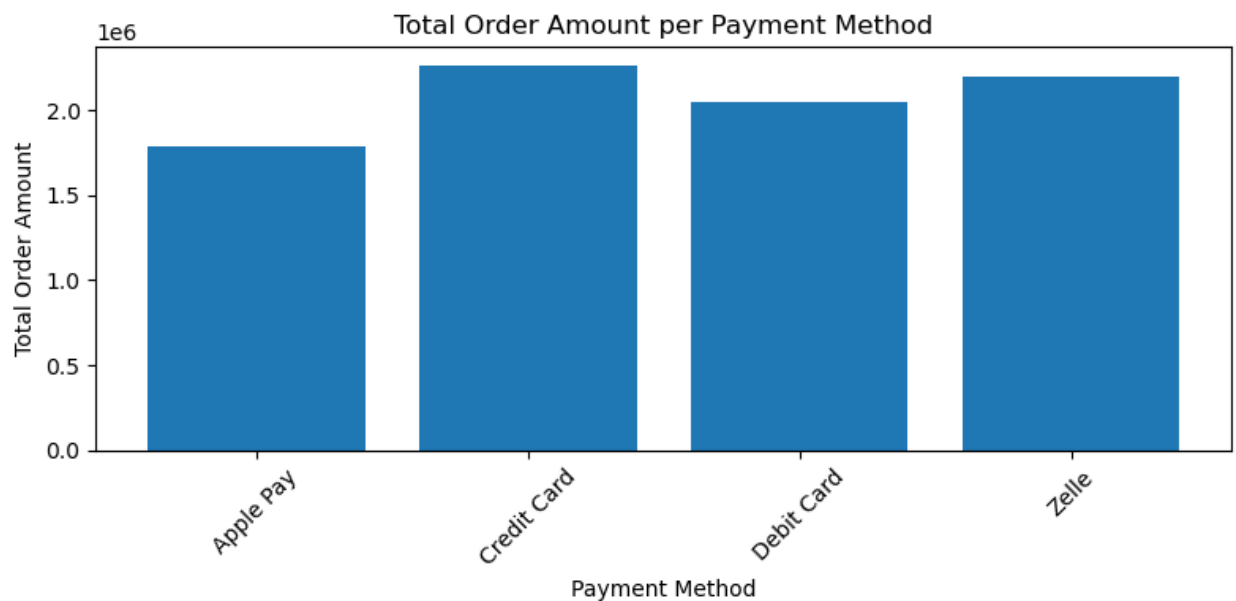
## Payment Analysis

We added a column Payment Method and mocked up following data 'Credit Card', 'Debit Card', 'Zelle' and 'Apple Pay'. We randomly distributed those values to all the customers.

- **Most Used Payment Method:**

Payment Method with the maximum count is 'Zelle'

- **Relationship between Payment Method and Order Amount:**



## Customer Behavior

- **Customer Activity Duration:**

On average, customers remain active for 133.34 days between their first and last purchase. This metric provides insights into customer loyalty.

- **Customer Segmentation:**

Based on purchase behavior, customers can be segmented into five groups:

- Top Customers
- High-Value Customers
- Medium-Value Customers
- Low-Value Customers
- Lost Customers

### Customer Segments:

	Recency	Frequency		Segment
12346	325	2	Low	Activity
12347	1	182	High	Activity
12348	74	31	Low	Activity
12349	18	73	High	Activity
12350	309	17	Low	Activity
...	...	...		...
18280	277	10	Low	Activity
18281	180	7	Low	Activity
18282	7	13	Low	Activity
18283	3	721	High	Activity
18287	42	70	High	Activity

Understanding these segments enables tailored marketing and retention strategies.

### Returns and Refunds

- Percentage of Orders with Returns or Refunds:

Approximately 2.21% of the orders have experienced returns or refunds. Monitoring and managing.

### Profitability Analysis

In order to calculate profit, we randomly distributed Unit Price values bit higher than the given Unit price value.

Post validations, we have got result sets as follows,

Total profit generated:

568057.0240000001

Top 5 products with higher profits:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Total_Spend	OrderDay	OrderTime	OrderHour	DayOfWeek
540421	581483	23843	PAPER CRAFT , LITTLE BIRDIE	80995	2011-12-09 09:15:00	2.08	16446	United Kingdom	168469.60	2011-12-09	09:15:00	9	4
61619	541431	23166	MEDIUM CERAMIC TOP STORAGE JAR	74215	2011-01-18 10:01:00	1.04	12346	United Kingdom	77183.60	2011-01-18	10:01:00	10	1
222690	556444	22502	PICNIC BASKET WICKER 60 PIECES	60	2011-06-10 15:28:00	649.50	15098	United Kingdom	38970.00	2011-06-10	15:28:00	15	4
173382	551697	POST	POSTAGE	1	2011-05-03 13:46:00	8142.75	16029	United Kingdom	8142.75	2011-05-03	13:46:00	13	1
348325	567423	23243	SET OF TEA COFFEE SUGAR TINS PANTRY	1412	2011-09-20 11:05:00	5.06	17450	United Kingdom	7144.72	2011-09-20	11:05:00	11	1



## Customer Satisfaction

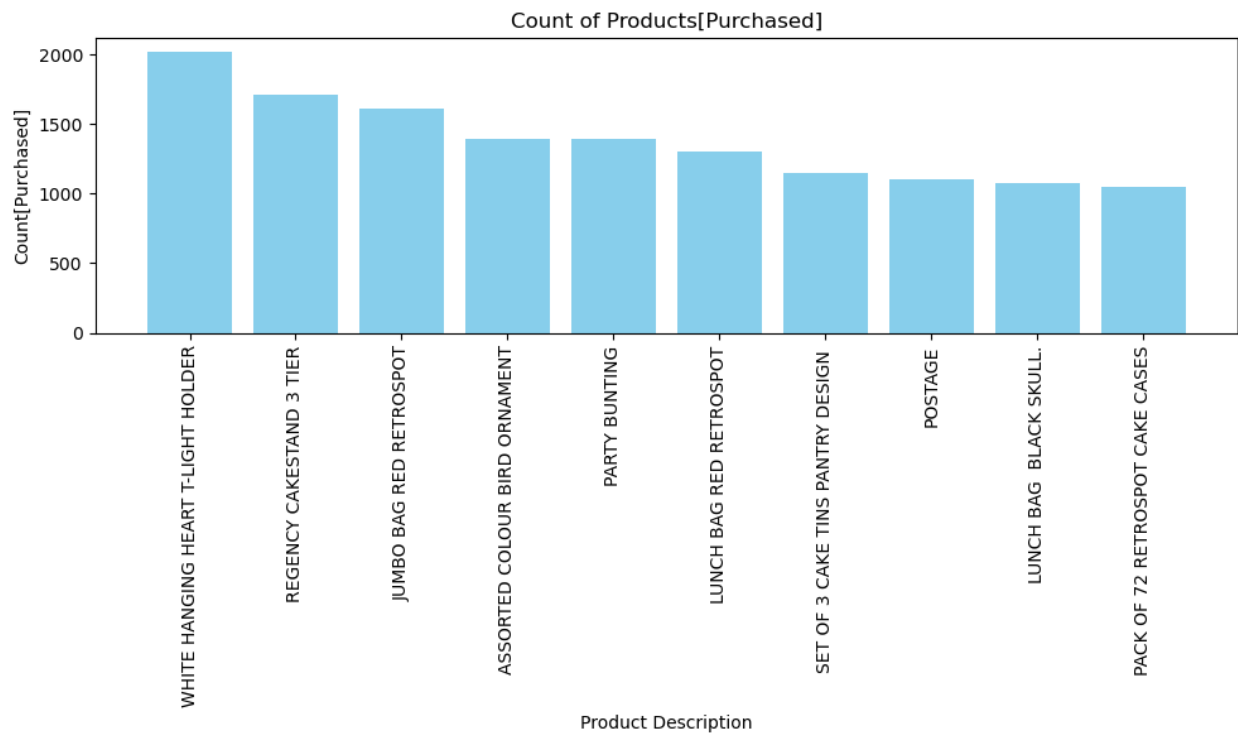
### Assumption:

- 1) Top most purchased products will fall under higher satisfactory category
- 2) Top most returned products will fall under dissatisfied category

### Results:

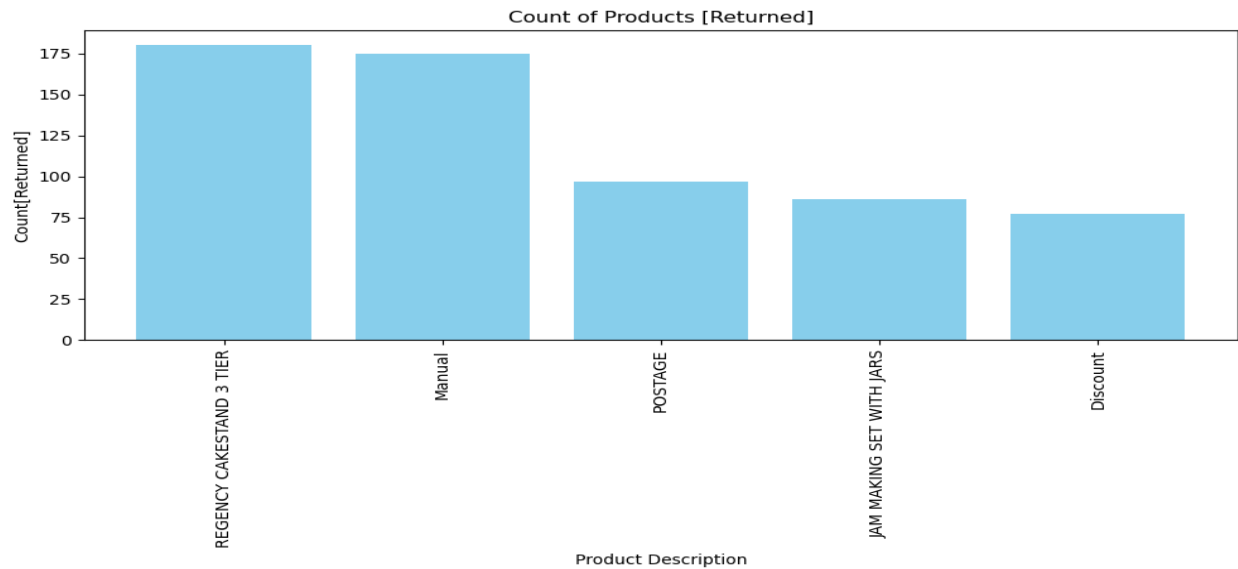
Most frequently bought products:

	Description	count
0	WHITE HANGING HEART T-LIGHT HOLDER	2016
1	REGENCY CAKESTAND 3 TIER	1714
2	JUMBO BAG RED RETROSPOT	1615
3	ASSORTED COLOUR BIRD ORNAMENT	1395
4	PARTY BUNTING	1390
5	LUNCH BAG RED RETROSPOT	1303
6	SET OF 3 CAKE TINS PANTRY DESIGN	1152
7	POSTAGE	1099
8	LUNCH BAG BLACK SKULL.	1078
9	PACK OF 72 RETROSPOT CAKE CASES	1050



Most frequently returned products:

	Description	count
0	REGENCY CAKESTAND 3 TIER	180
1	Manual	175
2	POSTAGE	97
3	JAM MAKING SET WITH JARS	86
4	Discount	77



## RFM Calculations:

We've also performed RFM calculations to gain insights into customer behavior:

- Days Since Last Purchase: This represents the number of days since a customer's most recent purchase.
- Total Products Purchased: The total number of products purchased by each customer.
- Total Transactions: The total number of transactions made by each customer.
- Total Spend: The total amount spent by each customer.
- Average Transaction Value: The average value of transactions for each customer.

Here is a snapshot of the RFM calculations for a sample of customers:

	CustomerID	RFM_Score
0	12346	1.70
1	12347	4.18
2	12348	3.01
3	12349	1.19
4	12350	0.68
...	...	...
4367	18280	0.70
4368	18281	0.79
4369	18282	2.97
4370	18283	4.73
4371	18287	2.71

These RFM metrics provide valuable insights into customer engagement, purchase frequency, and spending behavior, which can be used for customer segmentation and targeted marketing strategies.

This retail sales analysis report explores the dataset of a retail business, focusing on RFM (Recency, Frequency, Monetary) segmentation. The report involves data preprocessing, customer segmentation using K-means clustering, segment profiling, and marketing recommendations.

## RFM Segmentation

In this section, we applied RFM segmentation to the dataset to understand customer behavior and classify them into distinct segments.

### Recency Calculation

We calculated the recency of each customer, which represents the days since their last purchase. The dataset was transformed to include a recency score, which is an important indicator of customer activity.

### Frequency Calculation

The frequency score was determined by analyzing the total number of transactions per customer. Customers were ranked based on their transaction frequency and assigned a frequency score.

### Monetary Calculation

Monetary scores were assigned by examining the total spending per customer. The dataset now includes a monetary score, which reflects the customer's monetary contribution to the business.

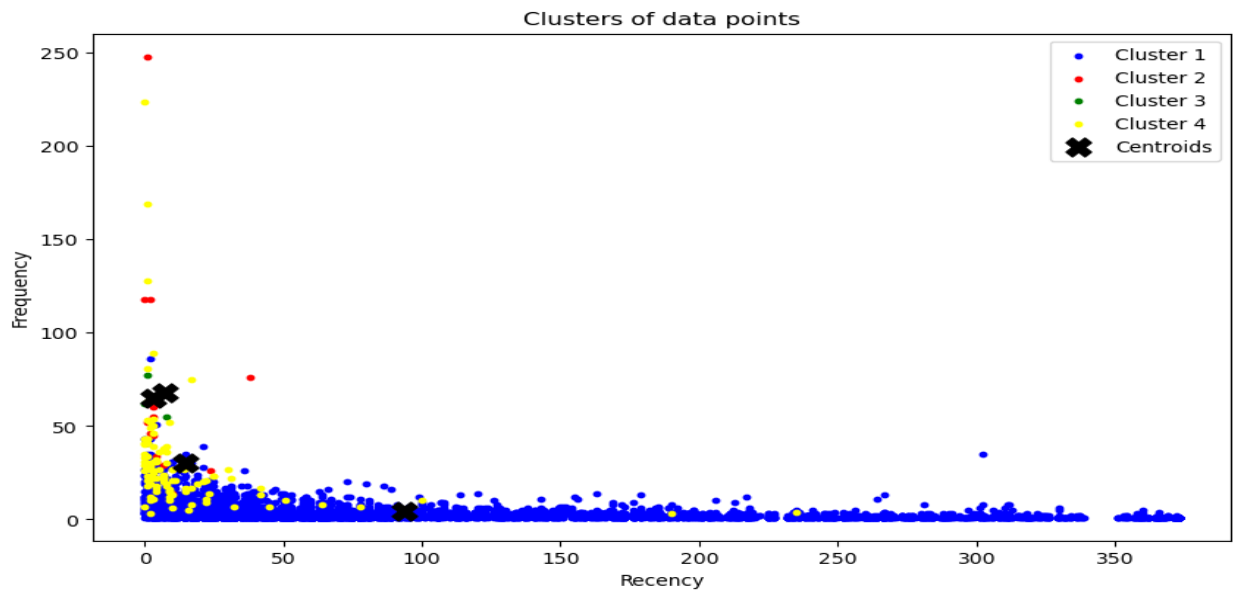
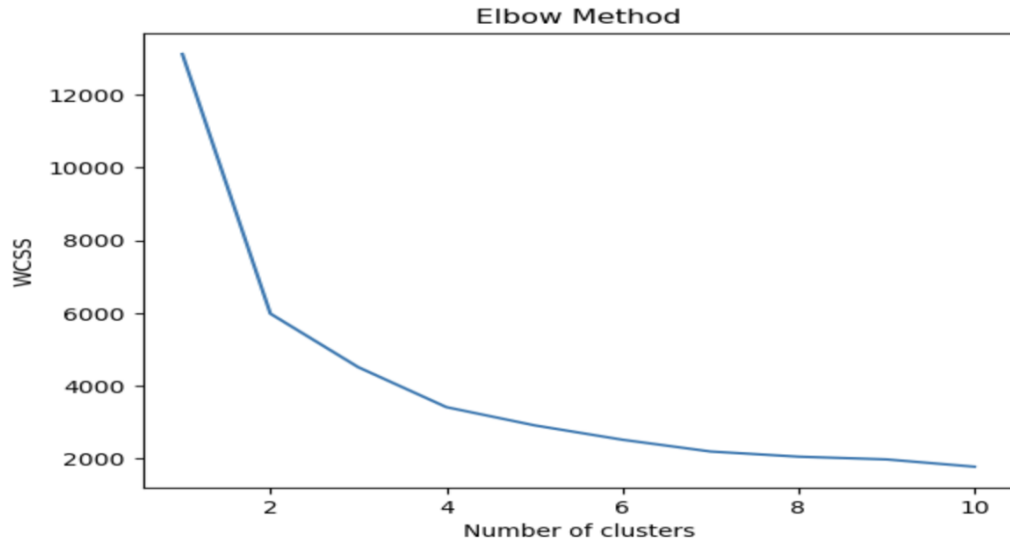
### RFM Score Creation

The Recency, Frequency, and Monetary scores were combined to create an RFM score for each customer. This score encapsulates customer behavior in one value.

	CustomerID	Recency	Frequency	Monetary	R_rank_norm	F_rank_norm	M_rank_norm	RFM_Score	Customer_segment
0	12346	325	2	0.00	3.88	39.39	39.39	1.70	Low Value Customers
1	12347	2	7	4310.00	95.81	81.43	81.43	4.18	High value Customer
2	12348	75	4	1797.24	37.73	64.25	64.25	3.01	Medium Value Customer
3	12349	18	1	1757.55	72.99	15.03	15.03	1.19	Lost Customers
4	12350	310	1	334.40	5.47	15.03	15.03	0.68	Lost Customers
...	...	...	...	...	...	...	...	...	...
4367	18280	277	1	180.60	8.38	15.03	15.03	0.70	Lost Customers
4368	18281	180	1	80.82	19.85	15.03	15.03	0.79	Lost Customers
4369	18282	7	3	176.60	87.95	54.33	54.33	2.97	Low Value Customers
4370	18283	3	16	2045.53	93.42	94.84	94.84	4.73	Top Customers
4371	18287	42	3	1837.28	53.96	54.33	54.33	2.71	Low Value Customers

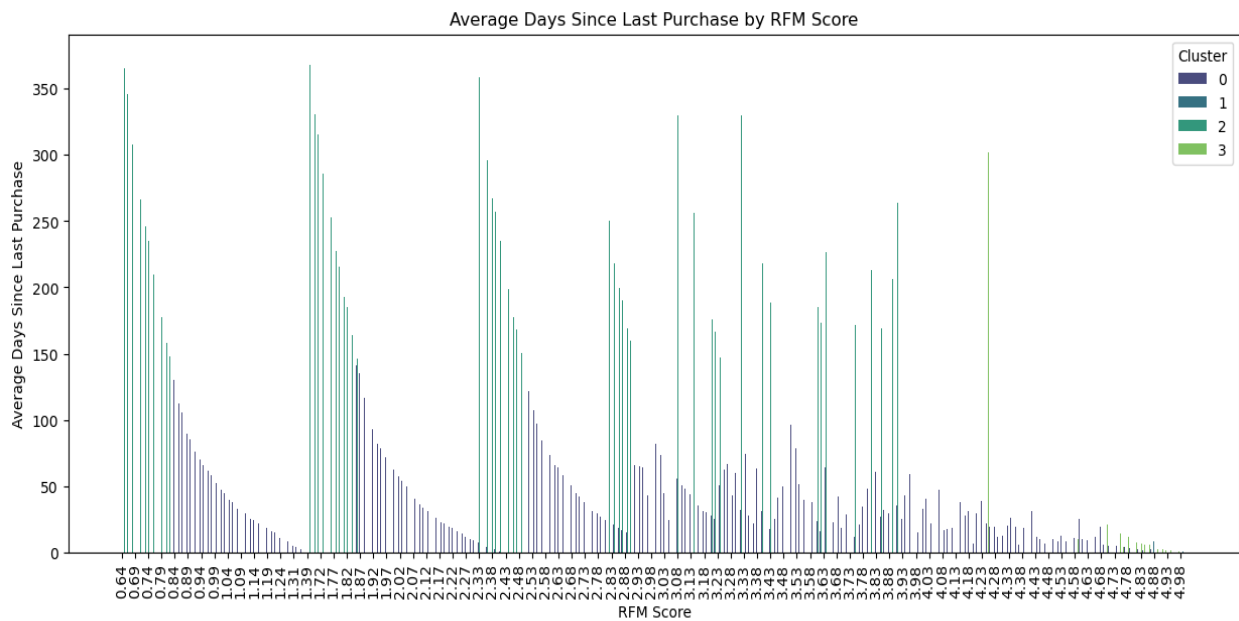
## Customer Segmentation

Customers were segmented into four categories based on their RFM scores. The segmentation provides insights into different groups of customers.



## Segment Profiling

This section provides an in-depth view of the segments created in the RFM analysis. It includes statistics for each segment, such as the average recency, products purchased, transactions, spending, and the average transaction value. Additionally, the distribution of customers across the segments is visualized using a bar plot.



## Marketing Recommendations

Tailored marketing recommendations were developed for each customer segment based on their characteristics and behavior. Recommendations aim to enhance customer engagement, retention, and spending. For example:

### Segment 0 Recommendations:

1. Offer an exclusive loyalty program to encourage purchases and increase frequency.
2. Send personalized product recommendations based on search patterns and previous purchases with special offers to re-engage customers.
3. Run email or SMS campaigns to keep customers connected with the brand.

### Segment 1 Recommendations:

1. Recognize and reward loyalty with member points or exclusive access to new products and events.
2. Suggest complementary products to increase the order value.
3. Ask for feedback on products and services to further tailor offerings.

### Segment 2 Recommendations:

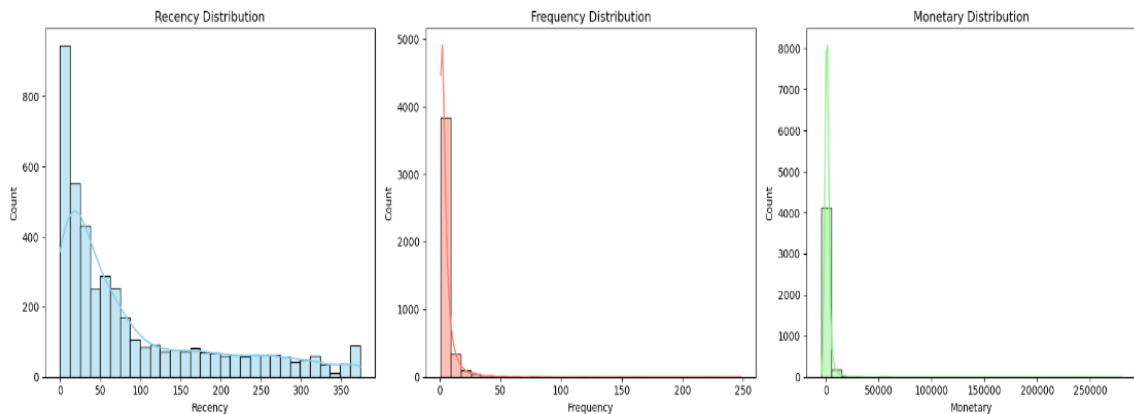
1. Recommend higher-value products or bundles to increase customer spending.
2. Create promotions for limited time offers to entice more frequent purchases.
3. Share content that highlights the benefits of your products to boost their perceived value.

### Segment 3 Recommendations:

1. Send targeted campaigns with attractive offers.
2. Identify the reasons for non-purchases and develop strategies to win customers back.
3. Ask for feedback and work on improving aspects that deterred customers from making purchases.

### Visualization

Visualizations were included to provide insights into the distribution of recency, frequency, and monetary metrics across the dataset. The graphs showcase the distribution characteristics and help in understanding customer behavior.



Out[99]:

	Cluster	Recency	Average_Products_Purchased	Average_Transactions	Average_Spend	Average_Transaction_Value	Customer_Count
0	0	41.144841	899.245188	4.802777	1472.916138	NaN	3169
1	1	7.166667	90192.500000	89.000000	182108.075000	2.240408	6
2	2	247.483901	256.295308	1.804968	451.036578	NaN	1087
3	3	8.609091	11121.772727	40.672727	18435.663364	2.000482	110

## Conclusion

This report delves into a comprehensive analysis of retail sales data, focusing on RFM (Recency, Frequency, Monetary) segmentation. This analysis offers valuable insights into customer behavior and provides a foundation for informed marketing strategies. The key findings and takeaways from the report can be summarized as follows:

1. **RFM Segmentation:** We successfully applied RFM segmentation to the dataset, which allowed us to classify customers into distinct groups based on recency, frequency, and monetary factors. This segmentation provides a clear understanding of customer engagement and purchasing patterns.
2. **Segment Profiling:** The report includes detailed segment profiles, highlighting the unique characteristics of each customer group. It is evident that there are substantial differences in recency, products purchased, transactions, and spending across the segments.
3. **Marketing Recommendations:** Tailored marketing recommendations were developed for each customer segment. These recommendations provide actionable strategies to enhance customer engagement and increase sales. The personalized approach to each segment ensures that marketing efforts are targeted and effective.
4. **Visualization:** Visualizations, such as histograms and bar plots, were used to represent the distribution of recency, frequency, and monetary metrics. These visual aids offer an easy-to-understand overview of customer behavior and provide insights for decision-makers.

In conclusion, the RFM segmentation analysis has provided valuable insights into the retail business's customer base. By understanding the different customer segments and their unique characteristics, the business can tailor its marketing efforts to better serve and engage customers. The recommendations provided in the report offer a clear roadmap for improving customer retention, increasing sales, and enhancing overall business performance. This analysis serves as a foundation for data-driven decision-making and marketing strategies in the retail sector.